

Trois issues qui doivent sommer à un

70,4 % de passages, 9,8 % de redoublements, 10,2 % d'abandons, 5,7 % d'achèvements et 2,9 % de transferts. Chacune exige le même dénominateur, et les transferts sont la raison pour laquelle ce n'est pas l'effectif inscrit.

Pierre Robentz Cassion

Leçon 5 sur 8

Analyse de programmes d'éducation — La cohorte

Ce que couvre cette leçon

- L'année, comme un ensemble de sorties
- Le dénominateur n'est pas l'effectif inscrit
- Le redoublement est une décision, non un événement
- L'abandon par année, et le pic de sixième
- Rapportez les sorties, puis les taux
- La suite

L'année, comme un ensemble de sorties — En Python

```
import pandas as pd

enrolment = pd.read_csv("school-enrolment-2024.v1.csv")
clean = enrolment[enrolment["age_years"] <= 20]
cohort = clean[clean["school_year"] == 2023]

outcomes = cohort["end_of_year_status"].value_counts(normalize=True)
print((outcomes * 100).round(1))
print(f"\ncohort: {len(cohort)} students")
```

L'année, comme un ensemble de sorties — En R

```
library(dplyr)

enrolment |>
  filter(age_years <= 20, school_year == 2023) |>
  count(end_of_year_status) |>
  mutate(share = n / sum(n))
```

L'année, comme un ensemble de sorties

Issue	Élèves	Part
Passés	910	70,4 %
Abandonné	132	10,2 %
Redoublé	127	9,8 %
Achevé la dernière année	74	5,7 %
Transféré	37	2,9 %
Encore inscrits à la clôture	12	0,9 %

L'année, comme un ensemble de sorties

- Les trois taux principaux sont le passage, le redoublement et l'abandon, et ils ne signifient quelque chose que s'ils. . .

Le dénominateur n'est pas l'effectif inscrit

- Les transferts ne sont pas des abandons — Un enfant parti dans une autre école est toujours scolarisé
- Les transferts ne sont pas non plus des réussites — Ils ne peuvent pas être comptés comme passages, car ce registre ne...

Le dénominateur n'est pas l'effectif inscrit — En Python (suite)

```
def rates(frame, exclude_transfers):
    base = frame[frame["end_of_year_status"] != "transferred-out"] \
        if exclude_transfers else frame
    return {
        "n": len(base),
        "promotion": (base["end_of_year_status"]
                      .isin(["promoted", "completed-final-grade"]).mean()),
        "repetition": base["end_of_year_status"].eq("repeated").mean(),
        "dropout": base["end_of_year_status"].eq("dropped-out").mean(),
    }

for exclude in (False, True):
    result = rates(cohort, exclude)
    label = "excluding transfers" if exclude else "all students"
    print(f"{label:20} n={result['n']} "
          f"promotion {result['promotion']:.1%} ")
```

Le dénominateur n'est pas l'effectif inscrit — En Python (suite)

```
f"repetition {result['repetition']:.1%}  "  
f"dropout {result['dropout']:.1%}")
```

Le dénominateur n'est pas l'effectif inscrit — En R

```
cohort |>
  filter(end_of_year_status != "transferred-out") |>
  summarise(n = n(),
            promotion = mean(end_of_year_status %in% c("promoted", "completed-final-grade")),
            repetition = mean(end_of_year_status == "repeated"),
            dropout = mean(end_of_year_status == "dropped-out"))
```

Le dénominateur n'est pas l'effectif inscrit

- C'est le dénominateur du taux de guérison PCIMA rencontré plus tôt dans ce module, dans un autre secteur — La...

Le redoublement est une décision, non un événement — En Python

```
by_district = clean.groupby("admin2")["repeating"].agg(["mean", "size"])  
print((by_district * [100, 1]).round(1))
```

Le redoublement est une décision, non un événement — En R

```
enrolment |> filter(age_years <= 20) |>  
  summarise(repeating = mean(repeating), n = n(), .by = admin2)
```

Le redoublement est une décision, non un événement

District	Taux de redoublement	n
Sud	12,3 %	583
Artibonite	12,1 %	618
Nord-Ouest	12,0 %	625
Centre	7,4 %	624

Le redoublement est une décision, non un événement — En Python

```
centre = clean[(clean["admin2"] == "Centre") & (clean["school_year"] == 2023)]
elsewhere = clean[(clean["admin2"] != "Centre") & (clean["school_year"] == 2023)]
for name, frame in [("Centre", centre), ("elsewhere", elsewhere)]:
    over = frame["age_years"] > frame["grade"] + 5
    print(f"{name}: repetition {frame['repeating'].mean():.1%}, "
          f"over-age {over.mean():.1%}")
```

Le redoublement est une décision, non un événement — En R

Repetition should predict over-age. Where it does not, suspect the coding.

Le redoublement est une décision, non un événement

- Un taux en décalage avec l'indicateur qu'il cause mécaniquement est une question sur le bureau avant d'être un constat. . .

L'abandon par année, et le pic de sixième — En Python

```
by_grade = cohort[cohort["grade"].between(1, 6)].groupby("grade").agg(  
    n=("student_id", "size"),  
    dropout=("end_of_year_status", lambda s: s.eq("dropped-out").mean()),  
    repetition=("end_of_year_status", lambda s: s.eq("repeated").mean()),  
)  
print((by_grade * [1, 100, 100]).round(1))
```

L'abandon par année, et le pic de sixième — En R

```
cohort |> filter(between(grade, 1, 6)) |>
  summarise(n = n(),
            dropout = mean(end_of_year_status == "dropped-out"),
            repetition = mean(end_of_year_status == "repeated"), .by = grade)
```

L'abandon par année, et le pic de sixième

Année	n	Abandon	Redoublement
1	149	8,1 %	2,0 %
2	271	8,5 %	7,0 %
3	332	13,0 %	6,6 %
4	231	8,7 %	10,4 %
5	165	8,5 %	19,4 %
6	110	18,2 %	12,7 %

L'abandon par année, et le pic de sixième

- Le redoublement monte avec l'année et l'abandon non — La cinquième retient 19,4 % de ses élèves alors que son abandon...
- L'abandon en sixième vaut 18,2 % sur 110 élèves — C'est le taux le plus élevé du tableau et le plus petit dénominateur,...

Rapportez les sorties, puis les taux — Exemple (suite)

End of school year 2023, 1,292 students

Promoted	70.4%	910	
Dropped out	10.2%	132	
Repeated	9.8%	127	
Completed the final grade	5.7%	74	
Transferred out	2.9%	37	excluded from the rates below
Still enrolled at cut-off	0.9%	12	

Rates on 1,255 students, transfers excluded:

Promotion (incl. completion)	78.4%
Repetition	10.1%
Dropout	10.5%

Centre reports repetition at 7.4% against 12.0-12.3% elsewhere while holding the highest over-age share of the four districts. Treat as a

Rapportez les sorties, puis les taux — Exemple (suite)

recording difference pending verification, not as performance.

- Ce sont les sorties d'une année.

Lire la leçon complète, avec le code exécutable

[https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/
edu-05-promotion-repetition-dropout](https://data-analysis.cassion.dev/fr/cours/education-programme-analysis/edu-05-promotion-repetition-dropout)